Для цитирования: Аналитика и контроль. 2018. Т. 22, № 3. С. 334-340
УДК 543.31 DOI: 10.15826/analitika.2018.22.3.001

Байесовский подход к повышению достоверности контроля качества вод

О.М. Розенталь[1], Л.Н. Александровская[2], А.В. Кириллин[2]

[1]Институт водных проблем РАН, Российская Федерация, 119991, г. Москва, ул. Губкина, д. 3
[2]Московский авиационный институт, Российская Федерация, 125993, г. Москва, А-80, ГСП-3, Волоколамское шоссе, д. 4.

**Адрес для переписки: Розенталь Олег Моисеевич, e-mail:* orosental@rambler.ru

Поступила в редакцию 14 мая 2018 г., после доработки – 19 июня 2018 г.

Повышенная изменчивость и одновременно – пониженная частота выборочных измерений контролируемых показателей природных вод повышают вероятность ошибочной оценки их качества. В работе решается задача повышения достоверности такой оценки путем анализа массивов новых данных совместно с данными, накопленными в предыдущие периоды. Для этого была применена модификация байесовского подхода с использованием показателя степени однородности объединяемых данных. Показано, что в последнем случае объединенная оценка смещается по сравнению с байесовской в сторону оценки максимального правдоподобия по вновь полученным экспериментальным данным, «забывая» таким образом устаревшие данные. При этом 90-процентный доверительный интервал, в котором заключены истинные значения контролируемых показателей, сужается, что повышает достоверность вероятностной оценки качества воды. Предложенный подход проиллюстрирован на примере универсального непараметрического метода оценки вероятности соответствия концентрации некоторого загрязняющего вещества предъявляемым требованиям, как наиболее общего показателя качества воды. Пример доведен до конкретных числовых значений, позволяющих как провести сравнение классического и модифицированного байесовского подхода, так и выдать рекомендации по рациональному использованию последнего. Предложенный подход может найти широкое применение в задачах анализа статистических показателей качества в различных предметных областях при дефиците экспериментальных данных.

Ключевые слова: контроль качества вод, вероятностная оценка, байесовский подход, смесь распределений, функция максимального правдоподобия

For citation: *Analitika i kontrol'* [Analytics and Control], 2018, vol. 22, no. 3, pp. 334-340
DOI: 10.15826/analitika.2018.22.3.001

Bayesian approach to improve the reliability of control of water quality

O.M. Rozental'[1], L.N. Aleksandrovskaya[2], A.V. Kirillin[2]

[1]Institute of water problems of RAS, ul. Gubkina, 3, Moscow, 125993, Russian Federation
[2]Moscow aviation Institute (MAI), Volokolamskoe shosse, 4, Moscow, 125080, Russian Federation

**Corresponding author: Oleg M. Rozental', e-mail: orosental@rambler.ru*

Submitted 14 May 2018, received in revised form – 19 June 2018

Increased variability and, at the same time, a reduced frequency of selective measurements of controlled indicators of natural waters increase the probability of erroneous evaluation of their quality. The task is to increase the reliability of such an assessment by analyzing arrays of new data in conjunction with data accumulated in previous periods. To do this, a Bayesian approach was modified using the uniformity measure of the combined data. It is shown that in the latter case the combined estimate shifts from the Bayesian one to the maximum likelihood estimate from the newly obtained experimental data, thus "forgetting" the obsolete data. At the same time, the 90% confidence interval, in which the true values of the monitored indicators are concluded, is narrowed, which increases the reliability of the probabilistic assessment of water quality. The proposed approach is illustrated by the example of a universal nonparametric method for

estimating the probability of the concentration of a certain pollutant in compliance with the requirements as the most common indicator of water quality. The example is brought to specific numerical values, allowing both to compare the classical and modified Bayesian approach, and to give recommendations on the rational use of the latter. The proposed approach can find wide application in the problems of analysis of statistical quality indicators in various subject areas with a shortage of experimental data.

Keywords: water quality control, probabilistic estimation, Bayesian approach, mixture of distributions, maximum likelihood function

Введение

Требования к достоверности контроля качества природных вод нарастают по мере того, как они аккумулируют все большее количество известных и новых загрязняющих веществ [1, 2]. Трудности решения этой задачи обусловлены сочетанием сравнительно редкого выборочного контроля с повышенной нестабильностью контролируемых показателей. Это создает повышенные риски ошибочных заключений о соответствии или несоответствии воды установленным требованиям [3, 4]. Для снижения таких рисков необходимо повысить объем контролируемой информации. Поскольку при этом нежелательно повышать частоту (а значит, и стоимость) измерений, общепринято стремление присоединять к свежим данным более старые, накопленных за предшествующий период. К сожалению, простое объединение не может быть эффективно, поскольку такие данные взаимно коррелированы тем меньше, чем больший интервал времени их разделяет. Такое объединение становится корректным, если учитывается факт снижения значимости старых данных в пользу более актуальных новых сведений. С этой целью необходима методика повышения достоверности контроля, в основу которой ниже положен байесовской подход.

1. Основные положения байесовского подхода

В модели Томаса Байеса совместная вероятность случайных событий A и B есть $P(A,B)=P(A\,|\,B)P(B)=P(B\,|\,A)P(A)$, откуда [5]:

$$P(B\,|\,A)=\frac{P(A\,|\,B)P(B)}{P(A)}, \qquad (1)$$

где $P(A\,|\,B)$ – функция правдоподобия, описывающая распределение случайной величины *A* при фиксированном значении B, $P(B\,|\,A)$ – апостериорная-расчетная плотность вероятности, $P(A)=\int P(A\,|\,B)P(B)dB$ – безусловная плотность вероятности (нормирующий множитель), Ω – область определения параметра *B* [5, 6].

Рассмотрим основные положения байесовского подхода на простейшем примере оценки вероятности непревышения некоторым загрязняющим веществом предельно допустимой концентрации (ПДК), принятой показателем качества воды. При этом шкалой измерений является наиболее «слабая» номинальная метрологическая шкала. Методы анализа результатов измерений, приведенных в этой шкале, относятся к классу универсальных непараметрических методов математической статистики. Они не требуют количественных значений данных, информации о законах распределения этих данных и оценок их характеристик (математических ожиданий, дисперсий и пр.). Эти методы являются альтернативой широко используемых в практических приложениях параметрических методов, основанных на предположениях о нормальном законе распределения вероятностей данных измерений, адекватность которого в качестве статистической модели результатов измерений концентрации загрязняющих веществ требуют специального исследования [3].

Процедура байесовского оценивания легко выполнима при экспериментальной оценке вероятности $\hat{R}=\frac{m}{n}$., когда из общего числа *n* наблюдений имеется $m \le n$ удовлетворительных и $d = n - m$ неудовлетворительных. Тогда функция правдоподобия представляет собой биномиальное распределение

$$P(m\,|\,n,R_{\text{П}})=\frac{n!}{n!(n-m)!}R_{\text{П}}^{m}\left(1-R_{\text{П}}\right)^{n-m}, \qquad (2)$$

где $R_{\text{П}}$ – вероятность удовлетворительного состава воды, индекс «П» здесь и далее характеризует правдоподобие величины.

Важно, чтобы априорная плотность вероятности была [7]: «самовоспроизводимой» (после коррекции по Байесу не должен меняться вид распределения), многообразной по форме (для адекватного описания априорных данных), «наихудшей» для гарантии точности получаемых апостериорных оценок (этому требованию отвечают распределения, полученные из условия максимальной неопределенности (энтропии)). Перечисленным условиям для функции (2) удовлетворяет бета-распределение:

$$P_0(R_0)=B(\gamma_0,\eta_0)R_0^{\gamma_0-1}\left(1-R_0\right)^{\eta_0-1}, \qquad (3)$$

где $B(\gamma_0,\eta_0)$ – бета-функция, $\gamma_0 \ge 1$, $\eta_0 \ge 1$ – ее параметры, индекс «0» здесь и далее относит величину к априорному распределению[1].

Выражения (2) и (3) позволяют при использовании формулы Байеса (1) получить выражение для апостериорной вероятности соответствия воды установленным требованиям [7]:

1 Подробное обоснование данного вида априорного распределения читатель найдет, например, в [6].

$$P(R_Б \mid m,n) = B(\gamma,\eta) R_Б^{\gamma-1}(1-R_Б)^{\eta-1}, \quad (4)$$

где новые параметры $\gamma = \gamma_0 + m$, $\eta = \eta_0 + n - m$, а индекс «Б» здесь и далее относит величину к ее значению по Байесу.

В качестве оценки параметра R выбираем апостериорное математическое ожидание

$$\hat{R}_Б = M[R_Б \mid m,n] = \frac{\gamma}{\gamma+\eta}, \quad (5)$$

где знак «^» над величиной здесь и далее характеризует оценку.

Если параметры γ_0, η_0 получены по результатам предшествующих наблюдений, то $\gamma_0 = m_0$, $\eta_0 = n_0 - m_0$. При этом, $\hat{R}_0 = \frac{m_0}{n_0}$ – оценка вероятности по частоте, найденная в предыдущий период. Для этого случая $\hat{R}_Б = \frac{m_0+m}{n_0+n}$,, т.е. байесовская оценка представляет собой оценку по объединенной выборке.

2. Ограничения байесовского подхода

Оценка $\hat{R}_Б$ является несмещенной только при условии однородности данных $M[\hat{R}_0] = M[\hat{R}_П] = M[\hat{R}_Б] = R$, критерий которой в задачах контроля при аппроксимации биномиального распределения нормальным имеет вид:

$$|u_0| = \frac{|\hat{R}_0 - \hat{R}_П|}{\sqrt{\frac{\hat{R}_0(1-\hat{R}_0)}{n_0} + \frac{\hat{R}_П(1-\hat{R}_П)}{n}}} \le u_{1-\alpha/2} \quad (6)$$

где $u_{1-\alpha/2}$ – квантиль стандартного нормального распределения уровня значимости 1-α/2.

Проверка статистической гипотезы об однородности связана с ошибками, и принимается с вероятностью 1 – α, т.е. только при нахождении статистики u_0 в допуске: $u_{\alpha/2} \le u_0 \le u_{1-\alpha/2}$. Действительная достоверность допускового контроля зависит от расстояния статистики u_0 от границы поля допуска, которое может быть измерено в вероятностной форме – наблюдаемом уровне значимости α_0, найденном из точного равенства $|u_0| = u_{1-\alpha_0/2}$. При условии $\hat{R}_0 = \hat{R}_П$ квантиль $u_{1-\alpha/2} = 0$ и $\alpha_0 = 1$; граница же поля допуска для принятия гипотезы однородности определяется на практике при значениях $\alpha_{гр} = [0.05 \div 0.1]$.

Для использования величины α_0 в качестве «меры» однородности/неоднородности можно произвести нормирование – ввести показатель $\alpha^* = \frac{\alpha_0 - \alpha_{гр}}{1-\alpha_{гр}}$, при $\alpha_0 = 1$ равный единице, а при $\alpha_0 = \alpha_{гр}$ – нулю. При промежуточных значениях α_0, например 0.8 и $\alpha_{гр} = 0.1$ получим $\alpha^* \approx 0.77$. При уменьшении $\alpha_{гр}$ имеем $\alpha^* \to \alpha_0$. Так, при $\alpha_{гр} = 0.01$ и $\alpha_0 = 0.8$: $\alpha^* = 0.798$; при $\alpha_{гр} = 0.001$ и $\alpha_0 = 0.8$: $\alpha^* = 0.799$. Следовательно, наблюдаемое значение α_0 также может быть использовано для оценки меры однородности как при принятии, так и при отклонении соответствующей гипотезы.

Пример 1. Проверить статистическую однородность данных по нефтепродуктам Уральского Управления по гидрометеорологии и мониторинга окружающей среды (УГМС) в 2007 и 2008 гг. в реке Исеть (левый приток р. Тобол) на створе в черте г. Екатеринбург, если из 12 результатов ежегодных измерений в 2007 г. было зафиксировано 11 удовлетворительных, а в 2008 – 6.

Решение. Исходные данные: 2007 г. – $n = 12$, $m = 11$, $\hat{R}_0 = \frac{11}{12} = 0.917$; 2008 г. – $n = 12$, $m = 6$, $\hat{R}_П = 0.5$.

Отсюда $u_0 = \frac{0.917 - 0.5}{\sqrt{\frac{0.917 \cdot 0.083 + 0.5 \cdot 0.5}{12}}} = 2.537 = u_{0.994}$.

Это значение больше квантиля стандартного нормального распределения $u_{1-\alpha/2} = 1.96$ при α = 0.05. Таким образом, данные контроля качества воды 2007 и 2008 гг. статистически разнородны. Они могут быть признаны однородными только при очень малом граничном значении $\alpha_{гр} = 0.01$.

Тогда нормированное значение при полученном выше $u_{0.994}$ есть

$$\alpha^* = \frac{2(1-0.994) - 0.01}{1-0.01} = \frac{0.012 - 0.01}{1-0.01} \approx 0.002.$$

Таким образом, объединять информацию, полученную в 2007-2008 гг. в соответствии с классическим байесовским подходом нельзя. Необходимо разработать подход, учитывающий степень статистической однородности, характеризуемой найденным значением показателя α*.

3. Обобщенная форма байесовских оценок.

Обобщенную форму байесовских оценок будем формировать на основе так называемой смеси распределений [7], имеющей два слагаемых, соответствующих байесовской и экспериментальной плотности вероятностей (функции правдоподобия):

$$P_{ОБ}(R) = \alpha^* P_Б(R) + (1-\alpha^*) P_П(R),$$

где индекс «ОБ» характеризует объединение выборок, с учётом степени статистической однородности.

При $\alpha^* = 1$ используется апостериорная плотность вероятности $P_Б(R)$; при $\alpha^* = 0$ происходит отказ от априорной информации и оценка строится на основе только функции правдоподобия $P_П(R)$ вновь поступивших данных.

Покажем процедуру построения байесовской оценки на рассмотренном выше примере нахождения вероятности по частоте.

Байесовская оценка имеет вид (5) $\hat{R}_Б = M[R_Б \mid m,n] = \frac{\gamma}{\gamma+\eta}$: $\hat{R}_Б = \frac{m_0+m}{n_0+n}$. Оценка максимального правдоподобия из (2) равна $\hat{R}_П = \frac{m}{n}$. С учетом выражения для обобщенной формы бай-

есовских оценок получим: $\hat{R}_{\text{ОБ}} = \alpha * \frac{m_0 + m}{n_0 + n} + (1 - \alpha *)\frac{m}{n}$, откуда, производя простые вычисления, имеем:

$$\hat{R}_{\text{ОБ}} = \frac{n_0 \alpha^*}{n_0 + n}\hat{R}_0 + (1 - \frac{n_0 \alpha *}{n_0 + n})\hat{R}_{\text{П}} \qquad (7)$$

Сравнивая это выражение с выражением байесовской оценки $\hat{R}_{\text{Б}} = \frac{m_0 + m}{n_0 + n} = \frac{n_0}{n_0 + n}\hat{R}_0 + \frac{n}{n_0 + n}\hat{R}_{\text{П}} = \frac{n_0}{n_0 + n}\hat{R}_0 + (1 - \frac{n_0}{n_0 + n})\hat{R}_{\text{П}}$, приходим к выводу о перераспределении суммарного объема выборки $n_0 + n$ в сторону уменьшения веса относительного априорного объема $\frac{n_0 \alpha^*}{n_0 + n}$ вместо $\frac{n_0}{n_0 + n}$ и увеличения веса относительного объема вновь поступившей информации о контролируемых показателях $\left(1 - \frac{n_0 \alpha^*}{n_0 + n}\right)$ вместо $\left(1 - \frac{n_0}{n_0 + n}\right)$.

4. Ожидаемая (прогнозируемая) оценка вероятности.

Ожидаемая оценка качества воды основана на так называемом предапостериорном анализе [5]. Таковым является анализ знаменателя формулы Байеса (1) апостериорного закона распределения вероятности *R* соответствия качества воды установленным требованиям. Этот знаменатель, равный $P(A) = \int_0^1 P(R_0)P(m \mid nR)dR$ представляет собой безусловный закон распределения ожидаемого числа *m* и при целых значениях γ_0, η_0 представляет собой бета-биномиальное распределение:

$$P(m) = \frac{m!(\gamma_0 - \eta_0 - 1)!(\gamma_0 + \text{m} - 1)!(\eta_0 + n - m - 1)!}{m!(n - m)!(\gamma_0 - 1)!(\eta_0 - 1)!(\gamma_0 + \eta_0 + n - 1)!}$$

с математическим ожиданием и дисперсией

$$M[m] = n\frac{\gamma_0}{\gamma_0 + \eta_0} = n\hat{R}_0,$$

$$D[m] = n(\gamma_0 + \eta_0 + n)\frac{\gamma_0 \eta_0}{(\gamma_0 + \eta_0)^2(\gamma_0 + \eta_0 + 1)} = n(\gamma_0 + \eta_0 + n)D[\hat{R}_0].$$

Отсюда ожидаемая оценка вероятности $\hat{R}_{\text{пр}}$ при проведении *n* «будущих» наблюдений равна оценке априорной

$$\hat{R}_{\text{пр}} = \frac{M[m]}{n} = \hat{R}_0, \qquad (8)$$

дисперсия которой повышена:

$$D\left[\hat{R}_{\text{пр}}\right] = D\left[\hat{R}_0\right] \cdot \frac{\gamma_0 + \eta_0 + n}{n}. \qquad (9)$$

Здесь $\gamma_0 + \eta_0 + n = n + n_0$ – количество наблюдений объединенной выборки, индекс «пр» характеризует вероятностное значение ожидаемой (прогнозируемой) величины.

На основе выражений (8), (9) с использованием нормальной аппроксимации может быть записан доверительный интервал, определяющий границы изменения оценки $\hat{R}_{\text{пр}}$:

$$\hat{R}_0 \pm u_{1-\alpha/2}\sigma_{\text{пр}}\left[\hat{R}_{\text{пр}}\right], \qquad (10)$$

где $u_{1-\alpha/2}$ – квантиль стандартного нормального распределения. При получении экспериментальной оценки $\hat{R}_{\text{П}} = \frac{m}{n}$ и сравнении ее с границами доверительного интервала $\underline{R}_{\text{пр}}$, $\overline{R}_{\text{пр}}$, могут быть сделаны следующие выводы:

1. $\underline{R}_{\text{пр}} \le \hat{R}_{\text{П}} \le \overline{R}_{\text{пр}}$ – экспериментальная оценка соответствует ожидаемой;

2. $\hat{R}_{\text{П}} < \underline{R}_{\text{пр}}$ – экспериментальная оценка ниже ожидаемой;

3. $\hat{R}_{\text{П}} > \overline{R}_{\text{пр}}$ – экспериментальная оценка выше ожидаемой.

Использование нормальной аппроксимации позволяет оценить также меру несоответствия экспериментальной и вероятностной оценки. Так, для наиболее важного водопользователю случая 2 ожидаемая степень снижения вероятности $\hat{R}_{\text{П}}$ по сравнению с $\hat{R}_{\text{пр}}$, т.е. $\Delta = \hat{R}_{\text{пр}} - \hat{R}_{\text{П}}$ с вероятностью β определится по соотношению [7]:

$$\frac{\Delta}{\sigma_{\text{пр}}} = \frac{u_{1-\alpha} - u_{\beta}}{\sqrt{n}}, \qquad (11)$$

где вследствие рассмотрения только нижней границы $u_{1-\alpha/2}$ заменено на $u_{1-\alpha}$, β – ошибка второго рода, определяющая ошибочную вероятность соответствия эксперимента и прогноза.

Анализируя теперь пошаговое изменение величины Δ в выборке нарастающего объема можно выявить закономерности динамики изменения состава воды.

Пример 2. Рассчитать прогнозируемую оценку вероятности соответствия по данным 2007 г., 2008 г. по условиям примера 1.

Решение. Ожидаемая оценка вероятности в 2008 г. совпадает с оценкой вероятности в 2007 г.

$$\hat{R}_{\text{пр.}} = \hat{R}_0 = 0.917,$$

однако имеет большую дисперсию

$$D\left[\hat{R}_{\text{пр.}}\right] = 0.1522,\ \sigma_{\text{пр.}}\left[\hat{R}_{\text{пр.}}\right] = 0.3901.$$

Нижняя односторонняя 90%-доверительная граница составляет $\underline{R}_{пр.} = 0.4178$.

Уменьшение оценки с вероятностью β может достигать в соответствии с (11) при α = β =0.05 величины Δ = 0.3724. Тогда ожидаемое значение искомой вероятности есть $\hat{R}_{\text{пр.}} - \Delta = 0.917 - 0.3729 = 0.5446$, что близко к экспериментальному значению $\hat{R}_{\text{П}} = 0.5$.

При необходимости объединения нескольких этапов наблюдений полученная оценка (7) рассматривается как априорная для последующего этапа и используется не только для оценивания искомого показателя, но и для проверки статистической однородности с последующим этапом. При этом удобно представить такую оценку в виде

$\hat{R}_{\text{ОБ}} = \frac{m_{\text{экв}}}{n_{\text{экв}}}$, где $n_{\text{экв}} = n_0 + n$, $m_{\text{экв}} = \text{I}\{R_{\text{экв}} \cdot n_{\text{экв}}\}$, знак I – целая часть числа.

Для трех этапов измерений имеем: n_0, m_0, $\hat{R}_0 = \frac{m_0}{n_0}$, α_1, n_1, m_1, $\hat{R}_{\text{П}_1} = \frac{m_1}{n_1}$; α_2, n_2, m_2, $\hat{R}_{\text{П}_2} = \frac{m_2}{n_2}$. Тогда в качестве априорной информации для третьего временного интервала рассматривается оценка (7) в виде: $\hat{R}_{\text{ОБ}_1} = \frac{n_0\alpha_1}{n_0 + n_1}\hat{R}_0 + \left(1 - \frac{n_0\alpha_1}{n_0 + n_1}\right)\hat{R}_{\text{П}_1}$, а для объединенной оценки $\hat{R}_{\text{ОБ}_2}$ получаем $\hat{R}_{\text{ОБ}_2} = \frac{(n_0 + n_1)\alpha_2}{n_0 + n_1 + n_2}\hat{R}_{\text{ОБ}_1} + \left[1 - \frac{(n_0 + n_1)\alpha_2}{n_0 + n_1 + n_2}\right]\hat{R}_{\text{П}_2}$. Подстановка в последнюю формулу выражения для $\hat{R}_{\text{ОБ}_1}$ и введение обозначений $\lambda_1 = \frac{n_0\alpha_1\alpha_2}{n_0 + n_1 + n_2}$; $\lambda_2 = 1 - \frac{(n_0 + n_1)\alpha_2}{n_0 + n_1 + n_2}$ приводит к искомой формуле:

$$\hat{R}_{\text{ОБ}_2} = \lambda_1\hat{R}_0 + (1 - \lambda_1 - \lambda_2)\hat{R}_{\text{П}_1} + \lambda_2\hat{R}_{\text{П}_2}. \qquad (12)$$

Сравнение выражений для $\hat{R}_{\text{ОБ}_1}$ и $\hat{R}_{\text{ОБ}_2}$ показывает, что веса λ_1 и $(1 - \lambda_1 - \lambda_2)$ более ранних оценок $\hat{R}_0$, $\hat{R}_{\text{П}_1}$ уменьшаются, т.е. происходит постепенное «забывание» предшествующей информации.

На практике чрезвычайно редок случай полного совпадения характеристик исследуемой информации, полученной в различные временные периода. Поэтому практически всегда α_1 и α_2 меньше единицы, и происходит приведенное выше перераспределение суммарного объема выборки в пользу увеличения веса вновь полученных данных, вне зависимости от того, «хуже» они или «лучше» предыдущих.

В результате объединенные оценки вероятности нахождения исследуемого показателя в допуске по частоте могут быть как хуже, так и лучше оценок максимального правдоподобия, полученных без учета ранее поступившей информации, в зависимости от того, уменьшается или увеличивается эта вероятность. Однако точность этих объединенных оценок, характеризующихся доверительным интервалом, всегда выше, чем точность оценок максимального правдоподобия. В то же время учет степени статистической однородности объединяемой информации приводит к некоторому ухудшению этих оценок и их точности по сравнению с байесовскими оценками, произведенными по суммарной выборке.

Пример 3. Применение модифицированной байесовской оценки вероятности соответствия качества воды предъявляемым требованям**.** Изложенный материал позволяет вернуться к приведенному во втором разделе статьи примеру оценки загрязнения воды нефтепродуктами с целью повышения достоверности анализа и вероятностной оценки качества воды путем учета вновь поступающей информации и своевременным отбрасыванием более ранних устаревших данных. Основой для этого будут служить байесовский метод оценивания и построение смеси распределений.

По данным примера 1 содержание нефтепродуктов в р. Исеть зафиксировано в 2007 г.: n_0 = 12, m_0 = 11; 2008 г. – n_1 = 12, m_1 = 6. При этом в 2009 г. там же – n_2 = 12, m_2 = 7.

Решение. Результаты расчётов, проведенных по формулам (5), (6), (7) и (12) сведены в таблицу.

Таким образом, использование априорной информации несколько повышает объединенную оценку. При этом по-прежнему имеет место выигрыш в точности. Действительно, видно, что 90-процентный доверительный интервал при оценке максимального правдоподобия есть $\hat{R}_{\text{П}_2} \pm u_{0.95}\sigma\left[\hat{R}_{\text{П}_2}\right]$ = 0.5833 ± 1.645·0.1423 = $(0.3492 \div 0.8174)$, а при объединенной оценке – $\hat{R}_{\text{ОБ}_2} \pm u_{0.95}\sigma\left[\hat{R}_{\text{ОБ}_2}\right]$ = 0.5474±1.645·0.0829 = = (0.4110÷0.6838), т.е. сужается. При этом нижняя, наиболее значимая для контроля водно-экологической безопасности, граница, при объединенной оценки почти на 20 % больше, чем при оценке максимального правдоподобия: $\frac{0.4110 - 0.3492}{0.3492} \approx 0.2$.

Предлагаемый метод может быть использован для анализа динамики качества воды не только во времени, но и в пространстве, например, на отдельных гидрохимических створах той же реки Исеть, где ряды данных не будут однородными.

Таблица

Результаты расчётов объединённых оценок

Table

Results of calculations of joint assessments

№ п/п	Объединяемые оценки	Оценки максимального правдоподобия (МП)	90 % доверительный интервал оценки МП	α*	Байесовская оценка	Объединённая оценка	90 % доверительный интервал объединённой оценки
1	2007 г.	$\hat{R}_0 = \hat{R}_{\text{П}_1} = 0.9167$	$[0.787; 0.9957]$	0.002	не возможна	0.5024	$[0.3346; 0.6702]$
2	2008 г.	$\hat{R}_{\text{П}_2} = 0.5$	$[0.2626; 0.7374]$				
3	2009 г.	$\hat{R}_{\text{П}_3} = 0.5833$	$[0.3492; 0.8174]$	0.6666	0.5277	0.5474	$[0.4110; 0.6838]$

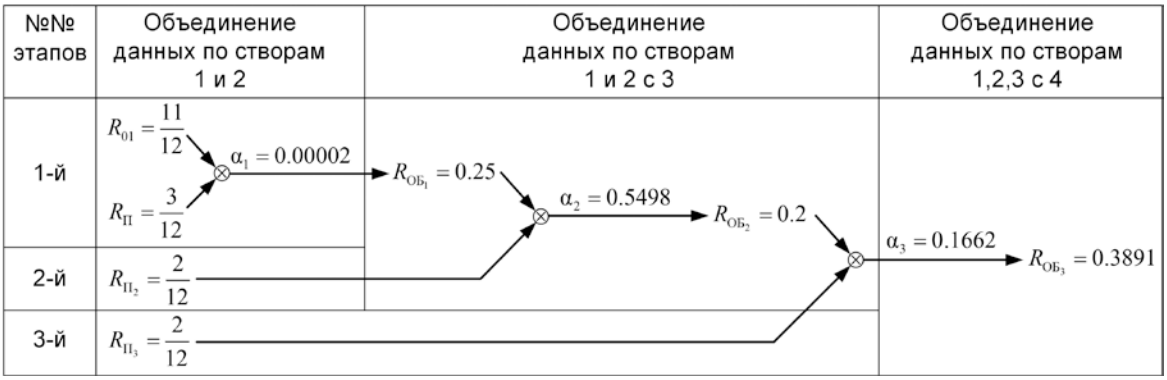

Рис. Структурная схема построения объединенной оценки концентрации нефтепродуктов в р. Исеть

Fig. Structural scheme for constructing a joint assessment of the concentration of petroleum products in the Iset' River

На рисунке схематически изображены основные результаты расчетов, проведенных по предложенной выше методике для анализа концентрации нефтепродуктов на створах р. Исеть в 2009 г. выше г. Екатеринбурга (створ 1), в черте города (2) и ниже на 7 (3) и 19.1 км (4).

При построении байесовской оценки с проверкой статистической однородности объединяемой информации, но без учета степени этой однородности, получили бы следующие результаты:

– данные, полученные на створе выше Екатеринбурга и в черте города статистически разнородны и не могут быть объединены;

– данные, полученные в черте города и ниже по течению однородны и могут быть объединены;

– результирующая оценка определяется по объединенной выборке. При этом, поскольку на створах 2,3,4, *m* = 3, 2 и 5 (2-й столбец рис. 1), имеем $R_{Б} = \frac{3+2+5}{36} = 0.2777$, что ниже, чем объединенная оценка $R_{ОБ_3} = 0.38$, полученная с учетом степени однородности. Таким образом, использование априорной информации целесообразно потому, что позволяет более объективно оценить исследуемые показатели.

Приведённый пример носит демонстрационный характер: проиллюстрировать целесообразность применения модифицированного байесовского подхода для итерационного (пошагового) мониторинга качества воды (в отличие от ступенчатого мониторинга по времени и пространству не учитывающего предысторию нестационарного процесса динамики качества воды). Поэтому авторы не ставили перед собой объёмную задачу полного анализа качества воду по ряду загрязняющих веществ.

Заключение

Достоверность вероятностных оценок качества воды может быть повышена путем объединения значений контролируемых показателей за несколько предыдущих лет. При этом необходим учет факта снижения значимости старых данных в пользу более актуальных новых сведений, для чего в работе предложено использование байесовского подхода и построение так называемой смеси распределений.

Показано, что байесовский подход как теоретическая основа для построения практических алгоритмов контроля динамики состава природных вод позволяет корректировать «свежими» данными более раннюю априорную информацию. Трудность такого подхода связана с тем, что указанную информацию необходимо описывать в виде некоторых функций распределения вероятностей, что даже в простейшем случае совместного оценивании математического ожидания и дисперсии нормального закона априорного распределения оказывается достаточно сложным. К тому же, если отсутствует однородность объединяемой информации, то байесовский подход не может применяться, т.к. дает смещенные результаты.

Разработан оригинальный метод, сочетающий в себе байесовское оценивание и формирование смеси распределений с весовыми коэффициентами, зависящими от степени статистической однородности объединяемой информации, как это продемонстрировано приведенным выше примером. Тем самым сформирована рабочая методика, апробированная на исследовании содержания нефтепродуктов в реке Исеть. Показано, что за счет придания большего веса «свежим» данным, объединенная оценка смещается по сравнению с байесовской в сторону оценки максимального правдоподобия. При этом 90-процентный доверительный интервал, в который заключены истинные значения контролируемых показателей сужается, что повышает вероятность правильного вывода о качестве воды.

ЛИТЕРАТУРА

1. Вода России. Речные бассейны: под науч. ред. А.М. Черняева; ФГУП РосНИИВХ. Екатеринбург: Издательство «Аква-Пресс», 2000. 536 с.

2. Zhen-Gang Ji. Hydrodynamics and water quality. Modeling rivers, lakes, and estuaries. A John Wiley & Sons, Inc, 2008. 670 p.

3. Александровская Л.Н., Розенталь О.М. Риск-ориентированный контроль содержания в воде загрязняющих веществ // Аналитика и контроль. 2016. Т. 20, № 1. С. 6-14.
4. Rene E.R., Saidutta M.B. Prediction of water quality indices by regression analysis and artificial neural networks // International Journal of Environmental Research. 2008. Vol. 2 (2). P. 183-188.
5. Gelman A., Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press, 2006. 648 p.
6. Carlin B.P., Louis T.A. Bayesian Methods for Data Analysis. Third Edition. Chapman & Hall/CRC, 2008. 535 p.
7. Методы анализа и оценивания рисков в задачах менеджмента безопасности сложных технических систем / Л.Н. Александровская [и др.]. СПб.: Корпорация «Аэрокосмическое оборудование», 2007. 462 с.

REFERENCES

1. Chernjaev A.M., eds. *Voda Rossii. Rechnye basseiny* [Water of Russia. River Basins]. Ekaterinburg, Aqua-Press Publ., 2000. 536 p. (in Russian).
2. Zhen-Gang Ji. *Hydrodynamics and water quality. Modeling rivers, lakes, and estuaries*. A John Wiley & Sons, Inc, 2008. 670 p.
3. Aleksandrovskaya L.N., Rozental O.M. [Risk-based monitoring of water pollutants]. *Analitika i kontrol'* [Analytics and Control], 2016, vol. 20, no. 1, pp. 6-14, DOI:10.15826/analitika.2015.20.1.004 (in Russian).
4. Rene E.R., Saidutta M.B. Prediction of water quality indices by regression analysis and artificial neural networks. *International Journal of Environmental Research*, 2008, vol. 2 (2), pp. 183-188, DOI: 10.22059/IJER.2010.192.
5. Gelman A., Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press, 2006. 648 p.
6. Carlin B.P., Louis T.A. *Bayesian Methods for Data Analysis. Third Edition*. Chapman & Hall/CRC, 2008. 535 p.
7. Kriukov S.P., Bodrunov S.D., Aleksandrovskaya L.N., Aronov I.E., Zakharevich A.P., Kuznetsov A.G., Kushel`man V.Ia. *Metody analiza i otsenivaniia riskov v zadachakh menedzhmenta bezopasnosti slozhnykh tekhnicheskikh sistem* [Methods for Analyzing and Estimating Risk in Security Management Problems for Complex Technical Systems]. Saint Petersburg, Aerokosm. Oborud. Corp. Publ., 2007. 462 p. (in Russian).